Preprints of the
Max Planck Institute for
Research on Collective Goods
Bonn 2013/12

The civil standard of proof – what is it, actually?

Mark Schweizer

MAX PLANCK SOCIETY

Preprints of the
Max Planck Institute
for Research on Collective Goods

Bonn 2013/12

The civil standard of proof – what is it, actually?

Mark Schweizer

July 2013

Max Planck Institute for Research on Collective Goods, Kurt-Schumacher-Str. 10, D-53113 Bonn
http://www.coll.mpg.de

The civil standard of proof – what is it, actually?

Mark Schweizer

Common Law distinguishes two standards of proof applicable in civil and criminal matters, respectively. The criminal standard of "beyond reasonable doubt" is much higher than the "preponderance of the evidence" standard used in civil cases. Continental European Civil Law, on the other hand, recognizes just one standard of "full conviction" applicable in both criminal and civil cases. This study is the first to look at the standard of proof actually used *by judges and judicial clerks in a Civil Law country (Switzerland). It is shown that, when asked directly, the members of court express a high decision threshold in line with legal doctrine and case law. But when Swiss judges are asked to estimate the error costs associated with each outcome and the error-cost-minimizing decision threshold is calculated based on the responses, the resulting standard is no different from the Common Law's "preponderance of the evidence" standard. When using the stated degree of belief in the truth of the plaintiff's allegations as a predictor for the grant of the plaintiff's request in a civil action, the probability of grant is 50% at a stated conviction of only 63%. It is further shown that the decision threshold is influenced by the individual's loss aversion, with individuals with a higher loss aversion having a higher decision threshold. No difference between the estimated decision threshold for members of the courts and members of the general population is found. The results suggest that the standard of proof actually employed by Swiss judges is not much different from the Common Law's "preponderance of the evidence" standard, despite the doctrinal insistence to the contrary.*

I. Introduction

Common Law knows (at least) two different standards of proof, the "preponderance of the evidence" (or "balance of probabilities" in English law) for civil cases and the "proof beyond reasonable doubt" in criminal cases.[1] In US law, a further intermediate standard of proof known as "clear and convincing evidence", which is applicable in certain civil cases (e.g., civil fraud), is well-established,[2] while it is a matter of controversy whether English law recognizes such an intermediate standard of proof.[3]

The "preponderance of the evidence" standard of US law is explained in the Federal Jury Practice and Instructions as follows:[4]

> *"To 'establish by a preponderance of the evidence' means to prove that something is more likely so than not so. In other words, a preponderance of the evidence in the case means such evidence as, when considered and compared with that opposed to it, has more convincing force, and produces in your minds belief that what is sought to be proved is more likely true than not true. This rule does not, of course, require proof to an absolute certainty, since proof to an absolute certainty is seldom possible in any case."*

For English law, a definition by Lord Denning in *Miller vs. Minister of Pensions* is the most cited paraphrase of the civil "balance of probabilities" standard. "If the evidence is such that the tribunal can say 'we think it more probable than not' then the burden is discharged, but if the probabilities are equal it is not."[5] If the fact finder is inclined to believe the plaintiff more than the defendant, even to the slightest degree, then he or she must find for the plaintiff.[6] In other words, it is sufficient if the plaintiff's allegations are more probably true than not.[7] On the other hand, to reach a guilty verdict in a criminal case, the jury must be convinced beyond any reasonable doubt that the facts alleged by the prosecution are true. A "reasonable" doubt is one that is

---

[1] See *Addington vs. Texas*, 441 U.S. 418 (1979), 422, 423; for English law *In Re H & Others* (minors) UKHL 16, AC 563 (1995), sect. 76; R. W. Wright, 'Proving Facts: Belief versus Probability', in H. Koziol and B. C. Steininger (eds.), *Tort and Insurance Law* (Vienna: Springer Vienna, 2009), 80.

[2] See, e.g, *Addington vs. Texas*, 441 U.S. 418 (1979), 422, 423.

[3] T. Anderson, D. A. Schum and W. L. Twining, *Analysis of evidence*, Law in context, 2.th edn. (Cambridge: Cambridge Univ. Press, 2006), 243; E. McBride, 'Is the civil 'higher standard of proof ' a coherent concept?' (2009) 8, Law, Probability & Risk, 323–51, 325 ssq.

[4] O'Malley K. F, Grenig J. E. and Lee W. C. (eds.), *Federal jury practice and instructions* (Eagan, MN: Westlaw, 2001), § 166.51.

[5] *Miller vs. Minister of Pensions,* 3 All ER 372 (1947), 373 sq.

[6] M. Redmayne, 'Standards of Proof in Civil Litigation' (1999) 62, Modern Law Review, 167–95, 172; for US law *Livanovitch v. Livanovitch*, 131 A. 799, 800 (Vt. 1926) ("If […] you are more inclined to believe from the evidence that he did so deliver the bonds to the defendant, even though your belief is only the slightest degree greater than that he did not, your verdict should be for the plaintiff" (quoting the jury instructions); Pennsylvania Suggested Standard Civil Jury Instructions, 3rd ed. 2005, § 1.42.

[7] llinois Supreme Court Committee on Pattern Jury Instructions in Civil Cases (eds.), Illinois Pattern Jury Instructions: Civil, § 21.01 ("more probably true than not true"), available from www.state.il.us/court/CircuitCourt/CivilJuryInstructions/21.00.pdf (last visited 1 August 2012); Sand L. B, Rakoff J. S, Reiss S. A, Loughlin W. P, Allen S. W. and Siffert J. S. (eds.), *Modern Federal Jury Instructions - Civil Volumes* (Los Angeles: LexisNexis, 2007), Bd. 4, § 73.01, Instruction 73–2 ("by a preponderance of the evidence" means "more likely true than not true").

based upon reason and not purely on speculation, a merely possible doubt does not prevent a finding against the defendant.[8]

Traditionally, the doctrinal discussion has emphasized that the standard of proof in civil matters is expressed by reference to the evidence or probabilities ("objectively"), while in criminal matters it is expressed by reference to the state of mind of the fact finder ("*convinced* beyond reasonable doubt").[9] However, as the jury instructions for the civil standard of the preponderance of the evidence – cited above – show, in civil cases, too, the belief or state of mind of the fact finder is what counts ("produces in your minds *belief* that what is sought to be proved is more likely true than not true"). The difference between the criminal and civil standard of proof lies in the *degree of belief* (or conviction) required for finding for the party bearing the burden of proof. This degree of belief is much higher in criminal cases than it is in civil cases.

It has long been suggested that normative decision theory provides an elegant explanation for the different standards:[10] since wrongly convicting an innocent person is widely considered to be a graver mistake than erroneously acquitting a guilty person, the expected error costs are minimized if the standard of proof in criminal cases is well above 50% (whether it can be quantified at all is highly controversial,[11] but nobody would dispute that a civil jury may find for the plaintiff under circumstances that would not permit a criminal jury to convict the accused). On the other hand, it is a commonly held assumption that in civil cases, the disutility of erroneously finding for or against the plaintiff is similar, which means the error-cost minimizing decision threshold is ≥50%.[12] As one commentator put it, "civil cases are the paradigm for symmetrical error

---

[8] Jury instructions according to the Ninth Circuit Model Criminal Jury Instructions, 2003 edition, § 3.5 – Reasonable Doubt – Defined. S. Sheppard, 'The Metamorphoses of Reasonable Doubt: How Changes in the Burden of Proof Have Weakened the Presumption of Innocence' (2003) 78, Notre Dame Law Review, 1165–250. shows the development from "moral certainty" to "reasonable doubt" to "articulate doubt", which, according to him, explains the current practice in criminal law better.

[9] Anderson, Schum and Twining, *Analysis of evidence,* 242.

[10] The seminal papers are J. Kaplan, 'Decision theory and the factfinding process' (1968) 20, Stanford Law Review, 1065–92 and A. D. Cullison, 'Probability Analysis of Judicial Fact-Finding: A Preliminary Outline of the Subjective Approach' (1969) 1, University of Toledo Law Review, 538–98.

[11] See the references cited in P. Tillers and J. Gottfried, 'Case comment--United States v. Copeland, 369 F. Supp. 2d 275 (E.D.N.Y. 2005): A Collateral Attack on the Legal Maxim That Proof Beyond A Reasonable Doubt Is Unquantifiable?' (2007) 5, Law, Probability & Risk, 135–57.

[12] *In re Winship* 397 U.S. 358, 371 (1970) (Harlan J. Concurring); V. Ball, 'The moment of truth: Probability theory and standards of proof' (1960) 14, Vanderbilt Law Review, 807–30, 817; D. H. Kaye, 'Apples and Oranges: Confidence Coefficients and the Burden of Persuasion' (1987) 73, Cornell Law Review, 4–77, 72; T. R. Lee, 'Pleading and Proof: The Economics of Legal Burdens' (1997), Brigham Young University Law Review, 1–34, 25; R. A. Posner, 'An Economic Approach to the Law of Evidence' (1999) 51, Stanford Law Review, 1477–546, 1504; Redmayne, 'Standards of Proof in Civil Litigation', 171; K. M. Clermont and E. Sherwin, 'A Comparative View of Standards of Proof' (2002) 50, American Journal of Comparative Law, 243–76, 268; A. Stein, *Foundations of evidence law* (Oxford, New York: Oxford Univ. Press, 2005), 148; E. Zamir and I. Ritov, 'Loss Aversion, Omission Bias, and the Burden of Proof in Civil Litigation' (2012) 41, Journal of Legal Studies, 165–207, 189; but see A. L. Tyree, 'Proof and probability in the Anglo-American legal system' (1982) 23, Jurimetrics, 89–100, 93 sq.

costs."[13] Therefore, "Bayesian decision theory seems to provide a pleasing and harmonious interpretation of civil litigation's usual requirement of proof by a preponderance of the evidence."[14]

Unlike Common Law, Continental European Civil Law does not generally distinguish between standards of proof for civil and criminal matters.[15] Standard of proof is always the (full) conviction of the judge, be it a "conviction intime" or a "conviction raisonnée", a reasoned or reasonable conviction (meaning that the judge must justify his or her decision by valid arguments).[16] This standard is described in the leading case of the German Federal Supreme Court (Bundesgerichtshof) as a "personal conviction […] in doubtful cases, the judge may and must be content with a degree of certainty useful for practical life that silences doubt without completely excluding it."[17] The Swiss Federal Supreme Court (Bundesgericht) uses a similar definition according to which "a court must be convinced of the truth of a factual allegation based on objective grounds. Absolute certainty is not required. It is sufficient if the court has no serious doubt or any remaining doubt appears insubstantial."[18] Neither German nor Swiss courts have ever expressed the decision threshold as a (quantified) subjective probability. The traditional doctrine is also reluctant to do so, but when it does quantify the standard of proof, the decision threshold is said to be above 90%;[19] sometimes figures of 95%[20] or even 99.8%[21] are given.

There are, certainly in German and Swiss law, many exceptions to the standard of full conviction in civil cases, namely for allegations that are notoriously difficult to prove, such as causality in medical malpractice or

---

[13] Lee, 'Pleading and Proof', 25.

[14] Kaye, 'Confidence Coefficients and the Burden of Persuasion', 55.

[15] C. Engel, 'Preponderance of the Evidence versus intime conviction: A Behavioral Perspective on a Conflict between American and Continental European Law' (2009) 33, Vermont Law Review, 435–67, 435; R. Motsch, 'Comparative and Analytical Remarks on Judicial Fact-Finding', in H. Rüßmann (ed.), *Festschrift für Gerhard Käfer* (Saarbrücken: Juris GmbH, 2009), 242.

[16] The standard of proof in Germany is better described as a "conviction raisonnée" rather than the French "conviction intime"; see G. Deppenkemper, *Beweiswürdigung als Mittel prozessualer Wahrheitserkenntnis: Eine dogmengeschichtliche Studie zu Freiheit, Grenzen und revisionsgerichtlicher Kontrolle tatrichterlicher Überzeugungsbildung (§ 261 StPO, § 286 ZPO)* (Göttingen: V&R Unipress, 2004), 208 sq, 421 and the references cited therein.

[17] BGHZ 53, 245 = BGH NJW 1970, 946 (translation from German by the author).

[18] BGE 130 III 321 sect. 3.2 (translation from German by the author).

[19] For German law T. Kadner Graziano, '"Alles oder nichts" oder anteilige Haftung bei Verursachungszweifeln?: Zur Haftung für perte d'unce chance/loss of chance und eine Alternative' (2011) 19, ZEuP, 171–200, 189; for Swiss law I. Berger-Steiner, *Das Beweismass im Privatrecht: Eine dogmatische Untersuchung mit Erkenntniswert für die Praxis und die Rechtsfigur der Wahrscheinlichkeitshaftung* (Bern: Stämpfli, 2008), sect. 6.81; H. P. Walter, 'Beweis und Beweislast im Haftpflichtprozess' (2009), Haftpflichtprozess, 47–68, 53; A. Bühler, 'Beweismass und Beweiswürdigung bei Gerichtsgutachten - unter Berücksichtigung der jüngsten Lehre und Rechtsprechung' (2010), Jusletter 21 June 2010, sect. 9.

[20] For German law R. Greger, *Beweis und Wahrscheinlichkeit: Das Beweiskriterium im Allgemeinen und bei den sogenannten Beweiserleichterungen* (Köln: Heymann, 1978), 110; for Swiss law D. Summermatter and C. Jacober, 'Zum Beweismass beim Kausal- und Motivationszusammenhang: Versuch einer begrifflichen Klärung anhand der jüngeren Rechtsprechung im Zivil-, Straf- und Sozialversicherungsrecht' (2012), HAVE, 136–49, 142.

[21] For German law R. Bender, 'Das Beweismaß', in W. Grunsky, R. Stürner, G. Walter and M. Wolf (eds.), *Festschrift für Fritz Baur* (Tübingen: Mohr Siebeck, 1981), 258; B. Fuchs, *Das Beweismaß im Arzthaftungsprozess* (Frankfurt am Main: Peter Lang, 2005), 80.

the theft of an insured item to be proven by the policy holder.[22] In summary proceedings, such as preliminary injunction proceedings, an even lower standard of "prima facie evidence" (Glaubhaftmachung) is used, which is met if the factual allegations supporting the claim are more probably true than not.[23] But the exceptions – and the considerable doctrinal effort required for their justification – prove the rule.[24] Few would dispute that the degree of conviction required for finding for the plaintiff in Civil Law is much higher than that required by the Common Law's "preponderance of the evidence" or "balance of probabilities" standard.[25] In Germany in the 1970s, a number of scholars influenced both by (possibly misunderstood) Scandinavian doctrine and the decision theoretic framework of *Kaplan* have suggested introducing a general standard of a "balance of probabilities" in civil matters.[26] These propositions have been met with almost visceral rejection and are today widely considered heterodoxy.[27] This scholarly debate is further evidence that, at least in theory, the civil standard of proof in Germany is indeed (much) higher than in US or English law. I am inclined to believe that this also holds for other Civil Law countries, but *Taruffo* does make the valid point that it is dangerous to generalize based on a few examples.[28] Since I am most familiar with German and Swiss law, I shall restrict myself to these jurisdictions.

The insistence of Civil Law that the standard of proof in civil cases is "full conviction", meaning "near certainty" (notwithstanding many exceptions), has left Common Law lawyers puzzled. In a strongly worded article, *Kevin M. Clermont* and *Emily Sherwin* "rudely wonder[ed] how civilians can be so wrong."[29] The article has met with an equally strongly worded rebuttal,[30] but also a renewed interest of civilians in standards of

---

[22] See, e.g, for German law BGH NJW 1995, 2169; NJW 2004, 777; for Swiss law BGE 130 III 321 sect. 3.3; 132 III 715 sect. 3.2.

[23] For German law H. Prütting, '§ 294', in T. Rauscher, P. Wax and J. Wenzel (eds.), *Münchener Kommentar zur Zivilprozessordnung*, 3rd edn. (München: Beck, 2008) note 24; for Swiss law Berger-Steiner, *Beweismass,* sect. 6.155.

[24] G. Walter, *Freie Beweiswürdigung* (Tübingen: Mohr Siebeck, 1979), 184.

[25] But see P. Gottwald, 'Das flexible Beweismaß im englischen und deutschen Zivilprozess', in P. Gottwald, E. Jayme and D. Schwab (eds.), *Festschrift für Dieter Henrich zum 70. Geburtstag, 1. Dezember 2000* (Bielefeld: Gieseking, 2000), 175; M. Brinkmann, *Das Beweismaß im Zivilprozess aus rechtsvergleichender Sicht* (Köln, Berlin, München: Heymann, 2005), 3, who argue against any difference in principle between the German and the Common Law's standard of proof in civil cases.

[26] G. Kegel, 'Der Individualanscheinsbeweis und die Verteilung der Beweislast nach überwiegender Wahrscheinlichkeit', in K. H. Biedenkopf, H. Coing and E.-J. Mestmäcker (eds.), *Das Unternehmen in der Rechtsordnung: Festgabe für Heinrich Kronstein aus Anlass seines 70. Geburtstages am 12. September 1967* (Karlsruhe: C.F. Müller, 1967), 335; B. M. Maassen, *Beweismaßprobleme im Schadenersatzprozess* (Köln: Carl Heymanns Verlag KG, 1975), 5 ssq.; R. Motsch, 'Vom Prozess als Beweis zum Überwiegensprinzip', in U. Klug, T. Ramm, F. Rittner and B. Schmiedel (eds.), *Gesetzgebungstheorie, juristische Logik, Zivil- und Prozessrecht: Gedächtnisschrift für Jürgen Rödig* (Berlin, New York: Springer, 1978), 335 ssq.; E. L. Nell, *Wahrscheinlichkeitsurteile in juristischen Entscheidungen* (Berlin: Duncker & Humblot, 1983), 211; more recently G. Wagner, 'Ökonomische Analyse der Rechtsmittel', in R. Bork, T. Eger and H.-B. Schäfer (eds.), *Ökonomische Analyse des Verfahrensrechts: Beiträge zum XI. Travemünder Symposium zur Ökonomischen Analyse des Rechts (26. bis 29. März 2008)* (Tübingen: Mohr Siebeck, 2009), 172 sq.

[27] H. Prütting, 'Beweislast und Beweismaß: Der Einfluss Leo Rosenbergs und Karl Heinz Schwabs auf die Entwicklung eines modernen Beweisrechts' (2010) 123, ZZP, 135–45, 142.

[28] M. Taruffo, 'Rethinking the Standards of Proof' (2003) 51, American Journal of Comparative Law, 659–77, 660.

[29] Clermont and Sherwin, 'A Comparative View of Standards of Proof', 244.

[30] Taruffo, 'Rethinking the Standards of Proof'.

proof.[31] *Clermont* and *Sherwin* have not, however, been convinced that their original analysis, which came to the conclusion that Civil Law values the perceived legitimacy of the court system higher than a rational approach to judicial decision making, was wrong.[32]

Scholars have wondered for a long time whether the Civilian judges actually adhere to the high standard proclaimed by case law and doctrine.[33] So far, no empirical study has looked into quantifying the standard of proof *actually employed* by judges in a Civil Law country. This paper reports the results of the first such study. It demonstrates that the standard of proof that Swiss judges and judicial clerks proclaim to adhere to when asked directly is much higher than the standard of proof that would result if the decision threshold was chosen to minimize the expected error costs of the decision *even when the error costs are obtained from the subjects*. It also estimates that there is a 50% probability that a request in a civil action will be granted when the judge is convinced only to a degree of 63% that the factual allegations supporting the claim are true. This decision threshold is no different from that estimated for a sample of the general German population. It is also comparable to the decision threshold of 70% estimated for Israeli trial lawyers by *Zamir* and *Ritov*, although Israel adheres to the Common Law's "preponderance of the evidence" standard in civil matters.[34] The results suggest that the standard of proof actually employed by courts in Switzerland, a traditional Civil Law country, is much lower than the standard proclaimed by the doctrine and case law.

Additionally, the study tests whether loss aversion leads to a higher decision threshold; a hypothesis advanced by *Zamir* and *Ritov* in a recent contribution.[35] Using a different methodology – i.e., measuring the individual loss aversion of each participant and using it as a predictor for the decision threshold –unlike *Zamir* and *Ritov*, this study finds an influence of loss aversion that goes in the expected direction.

The rest of this paper is structured as follows: First, the three different methods of measuring the standard of proof employed are described. Using different methods is important because different methods lead to different results. In Section III, the hypothesis regarding the influence of loss aversion on the standard of proof and the method of testing it are explained. Section IV describes the samples of Swiss judges and judicial clerks participating in the study, as well as the sample of the general German population used as a

---

[31] Engel, 'Preponderance of the Evidence versus intime conviction'.; Motsch, *Comparative and Analytical Remarks on Judicial Fact-Finding*.; P. Kinsch, 'Probabilité et certitude dans la preuve en justice', in Institut Grand-Ducal (ed.), *Actes de la Section des Sciences Morales et Politiques* (Luxemburg, 2009).

[32] K. M. Clermont, 'Standards of Proof Revisited' (2009) 33, Vanderbilt Law Review, 469–88.

[33] W. H. Rechberger, 'Maß für Maß im Zivilprozess?: Ein Beitrag zur Beweismaßdiskussion', in H. Prütting (ed.), *Festschrift für Gottfried Baumgärtel zum 70. Geburtstag* (Köln: Carl Heymanns Verlag KG, 1990), 490; M. Einmahl, 'Zeugenirrtum und Beweismaß im Zivilprozess: Eine Fallstudie am Beispiel des Verkehrsunfallprozesses' (2001), NJW, 469–75, 474 sq.; Clermont and Sherwin, 'A Comparative View of Standards of Proof', 261; Zamir and Ritov, 'Loss Aversion, Omission Bias, and the Burden of Proof in Civil Litigation', 172.

[34] Ibid, 177.

[35] Ibid.

comparison. The survey method and questionnaire used are exposed next before the results of the study are reported in Section V, followed by their discussion.

II. Measuring standards of proof

Several methods of measuring standards of proof as subjective probabilities are known.[36] The direct rating or self-report method consists of simply asking the subjects to quantify the minimal threshold they require for a guilty verdict (or a grant of the request) on a meaningful numerical scale.[37] Among the indirect methods, two approaches are distinguished: the decision-theoretic approach and the parallel-ranking or rank-order method.[38] It is known that the three methods do not lead to the same decision thresholds.[39]

A. Direct rating or self-report method

The direct rating or self-report method is useful for assessing whether different verbal definitions of the standard of proof, e.g., in jury instructions, are actually perceived as requiring different subjective probabilities of guilt for a conviction. The problem with using this approach with sophisticated subjects such as judges is that the judges know the theoretically required standard of proof, and they will likely give the answer that is expected and not the threshold actually used. Obtaining a decision threshold by the direct method is still useful for comparison with the decision thresholds obtained using other measurement methods, but it is not suitable to answer the question, relevant here, of what the *actual* standard of proof employed by judges is.

B. Decision theory-based method

The first of two indirect methods is called the decision theory-based method.[40] It obtains from the subjects the disutilities (or costs) required to parametrize the following inequation

$$\Pr(p) \geq \frac{(D_{fp}-D_{cn})}{(D_{fn}-D_{cp}) + (D_{fp}-D_{cn})} = \frac{1}{1+\left(\frac{D_{fn}-D_{cp}}{D_{fp}-D_{cn}}\right)}. \quad (1)$$

Pr(p) is the probability that the plaintiff's (or prosecution's) allegations are true. The expected costs are minimized when Pr(p) meets or exceeds the value calculated according to the above equation. D_{fp} is the disutility of a false positive decision, i.e., convicting an innocent person or granting an unfounded claim; D_{fn}

[36] See F. Dane, 'In search of reasonable doubt' (1985) 9, Law and Human Behavior, 141–58, 143 ssq.

[37] Ibid, 143; R. Hastie, 'Algebraic models of juror decision processes', in R. Hastie (ed.), *Inside the Juror: The Psychology of Juror Decision Making* (Cambridge: Cambridge Univ. Press, 1993), 101.

[38] Dane, 'In search of reasonable doubt', 143 sq.; Hastie, *Algebraic models of juror decision processes,* 103 sq.

[39] Ibid, 102 ssq.; M. K. Dhami, 'On measuring quantitative interpretations of reasonable doubt' (2008) 14, Journal of Experimental Psychology: Applied, 353–63.

[40] Terminology according to Hastie, *Algebraic models of juror decision processes,* 103.

the disutility of a false negative, i.e., acquitting a guilty person or denying a well-founded claim; D_{cp} is the disutility of a correct positive, i.e., convicting the guilty or granting a well-founded claim; and D_{cn} is the disutility of a correct negative, i.e., acquitting the innocent or denying an unfounded claim. The equation can be rewritten for utilities instead of disutilities, but most legal scholars work with disutilities or costs.[41]

Estimating the decision threshold using the (dis)utilities associated with each possible outcome from the subjects reliably leads to lower decision thresholds than those elicited using the direct method. Even in criminal cases, the decision thresholds calculated using the decision theory-based method often barely exceed 50%.[42] For my purposes, the problem with the decision theory-based method is that it only allows the estimation of a *normative* decision threshold, i.e., where the threshold would have to be in order to maximize utility (minimize disutility) given the subject's expressed (dis)utilities. But it again does not measure where the threshold *actually* is. Still, obtaining the parameters necessary to estimate the normative decision threshold from sitting judges is interesting in and of itself, because the judges may or may not share the common belief that in civil cases, the error costs of a false positive and a false negative decision are equal.

C. Parallel-ranking method

The parallel-ranking (also rank-order) method was first employed by *Simon* in 1970.[43] It was originally developed for an in-between subject design. Half the subjects express their belief in the guilt of the accused as a subjective probability, the other half only makes a dichotomous guilty-innocent verdict judgment. The subjective probabilities are ranked from highest to lowest, and the minimum threshold for a guilty verdict is determined by counting down the probability ratings to the rank number corresponding to the proportion of guilty verdicts obtained in the dichotomous decision condition (in the illustrative example given in Table 1, the minimum value obtained would be 70).

---

[41] Kaplan, 'Decision theory and the factfinding process', 1071 sq;. Cullison, 'Probability Analysis of Judicial Fact-Finding', 565; Maassen, *Beweismaßprobleme,* 7 sq.; R. O. Lempert, 'Modeling Relevance' (1977) 75, Michigan Law Review, 1021–57, 1032; Bender, *Beweismaß,* 247 ssq.; D. H. Kaye, 'The Limits of the Preponderance of the Evidence Standard: Justifiable Naked Statistical Evidence and Multiple Causation' (1982) 1982, Law & Social Inquiry, 487–516, 496; D. H. Kaye, 'Bayes, Burdens and Base Rates' (2000) 4, International Journal of Evidence and Proof, 260–7, 260; D. H. Kaye, 'The error of equal error rates' (2002) 1, Law, Probability & Risk, 3–8, 6; Nell, *Wahrscheinlichkeitsurteile,* 210; Motsch, *Vom Prozess als Beweis zum Überwiegensprinzip,* 335 sq.; M. Gräns, *Das Risiko materiell fehlerhafter Urteile: (Diss. Uppsala 1995)* (Berlin: Duncker & Humblot, 2002), 238 sq. and Zamir and Ritov, 'Loss Aversion, Omission Bias, and the Burden of Proof in Civil Litigation', 172, use disutilities, while L. H. Tribe, 'Trial by Mathematics: Precision and Ritual in the Legal Process' (1971) 84, Harvard Law Review, 1329–93, 1379 sq.; S. Nagel, D. Lamm and M. Neef, 'Decision theory and juror decision making', in B. D. Sales (ed.), *The trial process*, Perspectives in law and psychology (New York: Plenum Press, 1981), vol. 2, 355 sq.; E. Bourmistrov-Jüttner, *Subjektive Wahrscheinlichkeitstheorie und rationale Entscheidungstheorie in Anwendung auf die Rechtspraxis* (München, 1987), 291 ssq; A. Hoyer, 'Der Konflikt zwischen richterlicher Beweiswürdigungsfreiheit und dem Prinzip "in dubio pro reo"' (1993) 105, ZStW, 523–56, 541; Redmayne, 'Standards of Proof in Civil Litigation', 169 sq. and E. Lillquist, 'Recasting Reasonable Doubt: Decision Theory and the Virtues of Variability' (2002) 36, U.C. Davis Law Review, 85–197, 131, use utilities.

[42] See values reported in table 4.3 in Hastie, *Algebraic models of juror decision processes,* 105.

[43] R. J. Simon, '"Beyond a Reasonable Doubt": An Experimental Attempt at Quantification' (1970) 6, Journal of Applied Behavioral Science, 203–9.

Table 1: Illustration of the parallel-ranking method

guilty?	yes	yes	yes	yes	yes	yes	no	no	no
subj. probability	77	76	76	75	75	71	70	69	66
rank	1	2	3	4	5	6	7	8	9

The in-between subjects design has the disadvantage of low statistical power.[44] A within-subject design on the other hand is problematic because in some studies, order effects were found, i.e., subjects who first expressed a guilty verdict gave higher subjective probabilities than subjects who first gave their probability rating and then made the dichotomous choice.[45]

This study uses a within-subject design, but varies the order of judgment and rating to be able to control for order effects. The decision threshold is estimated using a binary logistic regression with the guilty (or rather, grant) verdict as dependent variable and the expressed subjective probability as independent variable. This allows estimating the degree of belief at which there is a 50% probability that a randomly chosen subject will convict the defendant in a criminal case (or grant the request in a civil case). Arguably, the 50% probability of grant thus computed conforms to the mid-point between the upper and lower threshold obtained using the parallel-ranking method, because at this point, the best one can say (based on the rank order method) is that there is a 50% probability that the request will be granted. Using a logistic regression instead of the parallel-ranking method allows for a richer, more psychologically plausible model.[46] The advantage of this method is that it estimates the degree of belief *actually required* for a probability of x% that the subject will grant the claim, which is the parameter of interest.

III. Loss aversion and standard of proof in civil cases

A. Loss aversion

A well-established finding from behavioural economics is that people show reference point-dependent valuations.[47] Most people are risk-averse when choosing between a sure gain and a positive gamble, but risk-seeking when choosing between a sure loss or a negative gamble.[48] *Kahneman* and *Tversky*'s descriptive theory of human decision making, Prospect Theory, therefore posits an S-shaped value function that is concave in

[44] D. B. Wright, K. A. Strubler and J. P. Vallano, 'Statistical techniques for juror and jury research' (2011) 16, Legal and Criminological Psychology, 90–125, 98 sq.

[45] R. J. Simon and L. Mahan, 'Quantifying Burdens of Proof: A View from the Bench, the Jury, and the Classroom' (1971) 5, Law & Society Review, 319–30, 322; but see Dane, 'In search of reasonable doubt', 149 sq, who did not find any order effect.

[46] Wright, Strubler and Vallano, 'Statistical techniques for juror and jury research', 98.

[47] A. Tversky and D. Kahneman, 'Loss Aversion in Riskless Choice: A Reference-Dependent Model' (1991) 106, Quarterly Journal of Economics, 1039–61.

[48] A. Tversky and D. Kahneman, 'Advances in prospect theory: Cumulative representation of uncertainty' (1992) 5, Journal of Risk and Uncertainty, 297–323, 306.

the domain of gains and convex in the domain of losses.[49] *Kahneman* and *Tversky* observed something else as well, though: losses loom larger than gains. "The aggravation that one experiences in losing a sum of money appears to be greater than the pleasure associated with gaining the same amount."[50] Most people find symmetric bets of the form (x, .50; -x, .50) unattractive. The aversiveness of symmetric bets generally increases with the size of the stake.[51] The value function of Prospect Theory is therefore steeper for losses than for gains. The ratio of G/L that makes an even chance to gain G or lose L just acceptable lies between about 2 and 2.5 for both risky and riskless choice involving monetary outcomes and consumption goods.[52] This means that people experience about twice the disutility for a loss than they experience utility for a corresponding gain.[53]

B. Reference-dependent valuation and civil litigation

Civil litigation provides a natural "frame" for outcomes.[54] Generally, the plaintiff frames the outcome of the litigation as a gain compared to the status quo ante trial. Conversely, the defendant sees the outcome as a loss.[55] Note that it is not always the plaintiff who disrupts the "status quo", which can be hard to define.[56] In some cases, the roles of the parties may be interchangeable, depending on who initiates the proceedings.[57] However, it is safe to say that in an overwhelming majority of cases, it is the plaintiff who asks the court to impose a change of the status quo, e.g., make the defendant pay, turn over possession of a good or stop behaving in a certain way. In all of these cases, the losing defendant will conceive compliance with the judgment as a loss.

It has been shown both in laboratory experiments and in field studies that (presumed) litigants indeed perceive the status quo ante trial as the relevant reference point and behave in accordance with the predictions of prospect theory.[58] *Zamir* and *Ritov* have hypothesized that loss aversion may also influence the

---

[49] D. Kahneman and A. Tversky, 'Prospect Theory: An Analysis of Decision under Risk' (1979) 47, Econometrica, 263–92.; Tversky and Kahneman, 'Advances in prospect theory'.

[50] Kahneman and Tversky, 'Prospect Theory', 279.

[51] Ibid, 279.

[52] Tversky and Kahneman, 'Loss Aversion in Riskless Choice', 154.

[53] Tversky and Kahneman, 'Advances in prospect theory', 59, suggest that the median of the empirically observed values for the difference in weight of gains and losses is about 2.25.

[54] J. J. Rachlinski, 'Gains, Losses and the Psychology of Litigation' (1996), Southern California Law Review, 113–85, 118.

[55] Ibid, 128.

[56] Clermont and Sherwin, 'A Comparative View of Standards of Proof', 268.

[57] Ibid, 268.

[58] R. Korobkin and C. Guthrie, 'Psychological Barriers to Litigation Settlement: An Experimental Approach' (1994) 93, Michigan Law Review, 107–92,133 ssq; L. Babcock, H. S. Farber, C. Fobian and E. Shafir, 'Forming Beliefs About Adjudicated Outcomes: Perceptions of Risk and Reservation Values' (1995) 15, International Review of Law and Economics, 289–303, 296 sq; Rachlinski, 'Gains, Losses and the Psychology of Litigation', 130 ssq.

decision threshold of the judge in a civil case.[59] This requires a further assumption, namely that the judge – who does not himself or herself gain or lose anything from his or her decision – vicariously experiences the gain or loss of the parties. This assumption is plausible, however, based on research that shows that judges are indeed influenced by the party's perspective despite having nothing at stake themselves. Judges are more likely to recommend settlement to a plaintiff than to a defendant, thereby showing the same pattern of risk aversion that the parties themselves show.[60]

Zamir and *Ritov* have used a scenario that only differs in whether the plaintiff pursues a declaratory action or an action for remedies. Since in the declaratory action scenario, granting the request does not entail a change in the status quo, the decision threshold should arguably be lower than in the action scenario if the judges take loss aversion into account. However, *Zamir* and *Ritov* could not find a difference in the decision threshold between the declaratory action and the action scenarios.[61] Interestingly, the decision threshold estimated using the parallel-ranking method was 70%,[62] which corresponds pretty well to the normative threshold when weighing the disutility of a loss about 2.25 as much as the corresponding utility of a gain, as suggested by prospect theory.

C. Hypotheses to be tested

Based on the previous research on different elicitation techniques for the standard of proof, I expect to observe a higher decision threshold when using the direct rating method than when using the decision theory-based method. I also expect that the actual decision threshold is lower than the one stated overtly, which may be subject to demand effects. This hypothesis is tested using a logistic regression with the decision threshold as dependent variable, which allows exploring where the actual decision threshold lies, rather than the one stated explicitly or the one calculated based on normative considerations.

Finally, if loss aversion had an influence on the required threshold for granting the plaintiff's request in a civil action, then those individuals with a higher G/L ratio, or a stronger loss aversion, should have a higher decision threshold. While on average the G/L ratio is about 2, there is considerable heterogeneity in loss aversion.[63] Using the G/L ratio as a predictor for the decision threshold in a logistic regression exploits the heterogeneity in loss aversion to test the hypothesis regarding the influence of loss aversion on the decision threshold in civil cases.

---

[59] Zamir and Ritov, 'Loss Aversion, Omission Bias, and the Burden of Proof in Civil Litigation', 172.

[60] C. Guthrie, J. J. Rachlinski and A. J. Wistrich, 'Inside the Judicial Mind' (2001) 86, Cornell Law Review, 777–830, 777.

[61] Zamir and Ritov, 'Loss Aversion, Omission Bias, and the Burden of Proof in Civil Litigation', 177 ssq.

[62] Ibid, 180.

[63] E. Fehr and L. Goette, 'Do Workers Work More if Wages Are High? Evidence from a Randomized Field Experiment' (2007) 97, American Economic Review, 298–317, 300; S. Gächter, E. J. Johnson and A. Herrmann, 'Individual-level loss aversion in riskless and risky choices' (2010), CeDEX discussion paper series, 4.

D. Measuring loss aversion

Testing the hypothesis regarding the influence of loss aversion on the decision threshold requires measuring the individual loss aversion of each subject. This was done using a simple choice between a six lotteries.[64] The subjects were shown the list of lotteries according to Table 2 and were asked which lotteries they would accept (play) or decline (not play). Although this task appears at first sight to measure risk aversion rather than loss aversion, it actually measures loss aversion.[65] Risk aversion (i.e., a concave utility of wealth function) in such small-stakes lotteries would imply absurd degrees of risk aversion in high-stake lotteries.[66]

Table 2: The lottery choice task

Please indicate for each toss of the coin whether you would like to play (accept) or not play (decline) the game.		
	accept	*decline*
If the coin turns up heads, then you lose €2; if the coin turns up tails, you win €6.	o	o
If the coin turns up heads, then you lose €3; if the coin turns up tails, you win €6.	o	o
If the coin turns up heads, then you lose €4; if the coin turns up tails, you win €6.	o	o
If the coin turns up heads, then you lose €5; if the coin turns up tails, you win €6.	o	o
If the coin turns up heads, then you lose €6; if the coin turns up tails, you win €6.	o	o
If the coin turns up heads, then you lose €7; if the coin turns up tails, you win €6.	o	o

Applying Prospect Theory allows determining the loss aversion implied by the subject's choice among the above lotteries. A decision maker will be indifferent between accepting and rejecting the lottery if $w^{+}(0.5)v(G) = w^{-}(0.5)\lambda^{risky}v(L)$, where L denotes the loss in a given lottery and G the gain; v(x) is the utility of the outcome $x \in \{G, L\}$, λ^{risky} denotes the coefficient of loss aversion in the risky choice task; and $w^{+}(0.5)$ and $w^{-}(0.5)$ denote the probability weights for the 0.5-chance of gaining G or losing L, respectively. If we assume that the same weighting function is used for gains and losses, $w^{+} = w^{-}$ as proposed by *Prelec*,[67] only the ratio $v(G)/v(L) = \lambda^{risky}$ defines the implied loss aversion in the lottery choice task. A frequent assumption on v(x) is linearity (v(x) = x) for small amounts, which gives us a very simple measure of loss aversion: $\lambda^{risky} = G/L$.[68] I.e., an individual that declines to play any of the lotteries has an implied loss aversion of > 3; an individual that accepts the first lottery and declines all the others one of 3 and so forth. Someone who accepts all the lotteries exhibits a loss *affection* of ≤ 0.87.

---

[64] As in Ibid, 8, adapted from Fehr and Goette, 'Do Workers Work More if Wages Are High? Evidence from a Randomized Field Experiment', 313.

[65] M. Rabin, 'Risk Aversion and Expected-utility Theory: A Calibration Theorem' (2000) 68, Econometrica, 1281–92, 1288; V. Köbberling and P. P. Wakker, 'An index of loss aversion' (2005) 122, Journal of Economic Theory, 119–31, 124; Fehr and Goette, 'Do Workers Work More if Wages Are High? Evidence from a Randomized Field Experiment', 313; Gächter, Johnson and Herrmann, 'Individual-level loss aversion in riskless and risky choices', 8.

[66] Rabin, 'Risk Aversion and Expected-utility Theory: A Calibration Theorem'.

[67] D. Prelec, 'The Probability Weighting Function' (1998) 66, Econometrica, 497–527, 503.

[68] Gächter, Johnson and Herrmann, 'Individual-level loss aversion in riskless and risky choices', 9.

The lottery choice task should ideally be incentivized. At least one of the lotteries should be drawn and actually played out, i.e., the subjects should face the possibility of real gains and losses.[69] Since the study reported here uses sitting judges as subjects, this was not possible. I cannot force a participating judge to pay me CHF 2 (in Switzerland, the lottery was played with Swiss Francs rather than Euros) if he or she losses the coin toss and it is administratively complex to pay out any gains when the subjects are not present in a laboratory. While *Holt* and *Laury* find differences in choices between hypothetical and real lotteries,[70] other authors were unable to find systematic differences.[71] Yet other studies find differences for hypothetical versus real gains, but not for hypothetical versus real losses.[72] It is undeniable that incentivized experiments are preferable. In the case at hand, however, the study could either be conducted with real judges or with real incentives. Both at the same time simply cannot be realized.

IV. Method and Participants

A. Online questionnaire for direct rating and estimation using binary logistic regression

For the direct rating method and to obtain the probability judgments and verdicts to estimate the decision threshold using a binary logistic regression, an online questionnaire was used. The participants first answered some demographic questions and then read a scenario that was adapted from the "loan" scenario used by *Zamir* and *Ritov*.[73] According to the "action" condition of the scenario, the plaintiff has allegedly given a long-time friend a loan of CHF 20,000 and requests that the court orders the defendant to pay him back the CHF 20,000. There is no written contract and no receipt; there is, however, some circumstantial evidence such as a deposit of CHF 20,000 into the defendant's bank account at the time he allegedly received the loan, and witnesses that report that there was talk about a loan. In the "negative declaratory action" condition of the same scenario, the plaintiff is the alleged debtor and requests a declaration that he owes nothing to the alleged creditor.

After having read the scenario, the participants answer three control questions to check whether they have understood the scenario and, importantly, to know who the plaintiff is (manipulation check). Participants who do not correctly answer the three control questions are excluded from further participation. Those who

---

[69] G. W. Harrison and E. E. Rutström, 'Risk aversion in the laboratory', in J. C. Cox and G. W. Harrison (eds.), *Risk aversion in experiments* (Bingley: JAI, 2008), 154.

[70] C. A. Holt and S. K. Laury, 'Risk Aversion and Incentive Effects' (2002) 92, American Economic Review, 1644–55.

[71] J. Beattie and G. Loomes, 'The Impact of Incentives Upon Risky Choice Experiments' (1997) 14, Journal of Risk and Uncertainty, 155-168.; but see for risk aversion C. F. Camerer and R. M. Hogarth, 'The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework' (1999) 19, Journal of Risk and Uncertainty, 7-42, 23.

[72] N. Etchart-Vincent and O. l'Haridon, 'Monetary incentives in the loss domain and behavior toward risk: An experimental comparison of three reward schemes including real losses' (2011) 42, Journal of Risk and Uncertainty, 61-83.

[73] Zamir and Ritov, 'Loss Aversion, Omission Bias, and the Burden of Proof in Civil Litigation', 200 sq. („Experiment 1: Loan"). An English translation of the German version used in this study is reported in Appendix I.

pass the control question then state their degree of belief in the truth of the plaintiff's allegation using a sliding scale, where 0% indicates "the plaintiff's story is certainly false" and "100%" means "the plaintiff's story is certainly true" (see Figure 1; a movable button appears after a participant clicks on the slider). Then the participants are asked whether they would grant or deny the claim. The order of these two questions was varied to allow controlling for order effects.

Figure 1: Slider used for rating degree of belief

The participants then choose which lotteries to accept in the lottery choice task explained above. Finally, the participants are given a choice of three verbal definitions of standards of proof used in civil proceedings in Switzerland, i.e., the default standard of "full conviction", the intermediate standard of "high probability" (used, for example, to establish causality) and the lower standard of "prima facie evidence" (Glaubhaftmachung) used primarily in summary proceedings, and asked to identify the standard of "full conviction" (the verbal definitions are those used by the Federal Supreme Court in its case law).[74] Afterwards, they are asked to express the degree of belief required by the default standard of full conviction on an identical slider as the one used to rate the belief in the plaintiff's story.

The link to the questionnaire was sent out by email in the early summer of 2012 by the High Courts of Berne and Zurich to all judges and judicial clerks (Gerichtsschreiber) in the respective Canton. Participation was voluntary, the recipients were encouraged to participate, but were not compensated for their participation. Answering the questionnaire takes 10 to 15 min. Judicial clerks are lawyers who help draft the opinions of the court, do research, and generally discuss the case with the judge. The tasks of judicial clerks in Switzerland are similar to those of judicial clerks in the US, but unlike in the US, employment as a judicial clerk in Switzerland is not necessarily temporary. While most move on to other jobs after a couple of years, some stay for much longer (a fraction becomes judges; almost all judges are former judicial clerks).

A total of 186 members of courts (88 judges and 98 judicial clerks) participated in the survey. 26 (14%) were excluded after incorrectly answering at least one of the control questions. Among those excluded are as many men as women, and their mean age does not differ significantly from that of the non-excluded participants. Of the 160 participants who completed the questionnaire, 76 are judges and 84 judicial clerks. Further descriptive statistics are reported in Table 2. A majority of judges (and a slight majority of judicial clerks)

[74] For the standard of full conviction, see BGE 130 III 321 sect. 3.2; for the intermediate standard see BGE 132 III 715 sect. 3.1; for the "prima facie" standard, see BGE 138 III 232 sect. 4.1.1.

works primarily on civil cases, the others mostly on criminal cases and a small minority on enforcement and bankruptcy cases.

Table 2: Descriptive statistics for the court member sample

	Men	Mean age (SD)	Mean work experience in years (SD)	Primarily working in civil law
Judges (N = 76)	47 (62%)	48.7 (8.3)	18.7 (8.9)	45 (59%)
Judicial clerks (N = 84)	40 (48%)	32.7 (7.3)	4.8 (5.9)	44 (52%)
Total	87 (54%)	40.2 (11.2)	11.6 (10.8)	89 (56%)

There are a total of approximately 343 judges and 476 judicial clerks in the Cantons Berne and Zurich.[75] Hence, roughly 22% of the judges and 18% of the judicial clerks participated in the study. 66% of all judges are male, while only 40% of all judicial clerks are male.[76] As can be seen from the data reported in Table 2, these proportions are roughly reflected in the sample, which indicates that the sample is quite representative. Given that one usually is elected as a judge between the age of 30 and 35 and retires at 65,[77] the mean age of the judges in the sample of 48.7 years seems quite representative of the population, too.

To allow a comparison of the court members with members of the general population, the same questionnaire was administered to a sample of the general German population between 18 and 60 years of age in February 2012. The participants were recruited by a commercial panel provider and remunerated for their participation with credit points redeemable for goods. The questionnaire deviated in two points from the one used for the court members. Firstly, the question on how to decide the case was re-phrased by adding the text in italics: "How would you decide this case *if you were acting as a judge*?". Secondly, the final questions regarding the definitions of the standard of proof and the expression of the normatively required decision threshold as a degree of belief were omitted.

A total of 247 participants (mean age 41 years, SD = 12.6; 49.7% men) completed the questionnaire. According to self-reports, 141 (53%) of the participants were employed, 11 (4%) employed in a managerial position, 19 (8%) self-employed, 26 (11%) students and 60 (24%) "other".

[75] Data for Zurich from the annual report (Rechenschaftsbericht) of the High Court of Zurich for 2011 (latest figures), available from www.gerichte-zh.ch/organisation/obergericht/rechenschaftsbericht.html (last visited 10 March 2013). Data for Berne from the human resource department of the High Court, Mrs Sonja Hartmann (on file with author).

[76] These figures are based on categorization by first name, which is not absolutely accurate, as some first names do not allow determination of the gender. However, in German, such names are very rare.

[77] In Switzerland, judges at lower courts are elected by popular vote, at higher courts they are usually appointed by parliament.

B. Paper and pencil questionnaire for decision theory-based method

For the estimation of the error costs associated with each outcome of a civil action, a thought experiment inspired by *Laudan* and *Saunders* was implemented using a simple one page paper and pencil questionnaire.[78] The participants were attending a seminar on case management in civil proceedings organized by the High Court of Berne on 23 August 2012. All 49 attendees completed the simple questionnaire, which consisted of the scenario described in the next paragraph and some demographic questions on the back of a single sheet of paper. Completion of the questionnaire took less than 10 minutes; during this time, the participants were not observed to be talking to each other.

The participants were told that an anonymous donor had given them CHF 100,000 under the condition that they must spend the entire amount, but not more than the entire amount, on preventing the possible outcomes of the following simple case: An individual brought a civil action against another individual before a competent court, requesting the payment of CHF 100,000. Both parties have similar wealth and income. *The only issue is an issue of fact, namely, whether the plaintiff actually gave the defendant a loan in the amount of CHF 100,000.* It is undisputed that the loan, if it was granted, had not been paid back despite being overdue.

The four possible outcomes of the case are (italics in the original. The order of the outcomes was randomized):

a) The plaintiff *has* given the defendant a loan of CHF 100,000, and the defendant *is* ordered to pay CHF 100,000 to the plaintiff;

b) the plaintiff has *not* given the defendant a loan of CHF 100,000, but the defendant *is* ordered to pay CHF 100,000 to the plaintiff;

c) the plaintiff has *not* given the defendant a loan of CHF 100,000, and the defendant is *not* ordered to pay CHF 100,000 to the plaintiff;

d) the plaintiff *has* given the defendant a loan of CHF 100,000, but the defendant is *not* ordered to pay CHF 100,000 to the plaintiff.

Because the total amount to be invested in preventing the outcomes is fixed, the scenario avoids the difficult question of estimating the absolute error costs of each outcome and instead allows identifying the (only relevant) *ratio* of the error costs.

Two participants are excluded from the following analysis because they invested more than the donated amount. Of the 47 participants who correctly completed the questionnaire, 20 (43%) were judges, 10 in charge of conducting the (in most cases) mandatory conciliation hearing and 17 (36%) judicial clerks. The

[78] See L. Laudan and H. D. Saunders, 'Re-Thinking the Criminal Standard of Proof: Seeking Consensus about the Utilities of Trial Outcomes' (2009) 7, International Commentary on Evidence, 1–36, 23.

persons in charge of conducting the mandatory conciliation are fully qualified lawyers. Under Swiss civil procedural law, in most cases, a party can only bring suit after an attempt of settlement has failed, which must be conducted before the competent authority (see art. 202 ssq. Swiss Civil Procedure Act). At 57%, women were in the majority overall, but they were, unsurprisingly, a minority of 35% among the judges in the sample. Participants were between 29 and 61 years old, with a mean age of 40, the mean age of judges (47) being higher than that of the judicial clerks (32) and the persons in charge of reconciliation (38).[79] The participants had an average of 7.3 years of work experience; judges a mean of 12 years and judicial clerks a mean of 4 years. 13 of the participants had also participated in the online questionnaire (about two months earlier). For the reasons outlined above, the demographics of the sample are quite representative for the population.

V. Results

A. Results from the direct rating / self-report method

118 (73.8%) court members identified the correct verbal definition of the default standard of proof of "full conviction". 40 (25%) chose the verbal definition of the intermediate standard of "high probability", and only 2 the definition of "prima facie evidence". No significant differences in the proportion of correct answers of judges versus judicial clerks, men versus women and those who self-identified as working primarily in civil law versus those who work in other areas of law were found.

19 participants indicated that a 100% certainty was required under the standard of full conviction; the lowest threshold indicated was 51%. The median decision threshold under the standard of full conviction was 91% (M = 88.8, SD = 11.9). No difference in the mean decision threshold between those who correctly identified the verbal definition of the standard of full conviction (N = 118, M = 87.5, SD = 12.9) and those who incorrectly identified the verbal definition (N = 42, M = 90.4, SD = 8.3) was observed. No difference in the required decision threshold between those who granted the request (N = 69, M = 87.4, SD = 11.3) and those who denied the request (N = 91, M = 89.9, SD = 12.4) was found, either.

B. Results from the decision theory-based method

Table 3 reports the descriptive statistics for the apportionment of the CHF 100,000 among the four possible outcomes of the civil action.

[79] Nine participants did not state their age.

Table 3: Mean apportionment of CHF 100,000 to prevent outcomes (in brackets: median)

	loan was given	loan was not given
request is granted	CHF 2,979 (CHF 0)	CHF 47,766 (CHF 50'000)
request is denied	CHF 46,276 (CHF 50'000)	CHF 2,979 (CHF 0)

The modal answer, given by 34 (72%) of the participants, was to invest CHF 50,000 each in the prevention of a false positive (upper right cell in Table 3) and a false negative (lower left cell in Table 3) and nothing in the prevention of the correct decisions. The second most common answer, given by 3 participants, was to invest CHF 25,000 in the prevention of each outcome.

Taking the means from Table 3 and calculating the decision threshold using equation (1) results in a decision threshold of 51% (rounded). Calculating the decision threshold in this way depends on the (strong) assumption that each additional Swiss Franc invested in the prevention of an outcome has the same effect as the last. However, given that (almost) the same amounts were invested in the prevention of each of the correct and each of the incorrect decisions, the assumption hardly matters.

10 (21%) participants have invested different amounts in the prevention of the two types of errors. Seven of those invested more in the prevention of a false positive (grant of request despite no loan given) than in the prevention of a false negative (denial of request although loan was given). The decision thresholds resulting from equation (1) using the implied error costs indicated by these seven participants result with a single exception in thresholds below 63%. The implied error costs of just one participant result in a decision threshold of 87.5%, close to the mean threshold obtained with the direct rating method.

C. Results from the binary logistic regression

Excluded from the following analyses were the 24 participants (4 court members and 20 members of the general population) that granted the request although they indicated that their degree of belief in the truth of the allegations supporting the claim was less than 50% or denied the request although they indicated that they were 100% certain that the allegations supporting the claim were true.[80]

Table 4 reports the descriptive statistics from the online questionnaire. The first four columns report the results for the "action" condition, the last four columns those for the "negative declaratory action" condition (since the order of questions did not have an effect, as will be shown below, the results are pooled across the "order" condition).

[80] The results including those 24 subjects are reported in the Appendix II. All effects remain significant with the exception of type of request in Model 3. Model fit as measured by pseudo R^2 decreases with the additional 24 subjects.

Table 4: Descriptive statistics for the degree of belief and decision

	Action				negative declaratory action			
	grant		denial		grant		denial	
	obs. (share)	Ø conviction (SD)	obs. (share)	Ø conviction (SD)	obs. (share)	Ø conviction (SD)	obs. (share)	Ø conviction (SD)
members of court (N = 156)	62 (71%)	84.1 (10.2)	25 (29%)	57.7 (19.0)	3 (4%)	81.3 (11.1)	66 (96%)	25.4 (23.0)
general population (N = 227)	93 (58%)	79.9 (13.9)	68 (42%)	49.2 (23.5)	13 (20%)	63.5 (16.3)	53 (80%)	36.8 (26.2)

Unsurprisingly, the mean conviction of those who granted the request is higher than the mean conviction of those who denied the request for both conditions and both samples. A higher percentage of court members grants the claim than members of the general population (71% versus 58%; χ^2 [3 df, N = 383] = 21.323, $p < 0.001$). Very few members of court granted the negative declaratory action.

Looking at the mean conviction irrespective of whether the request was granted or denied, the allegations of the plaintiff in the "action" condition (the alleged creditor) were believed with a degree of belief of 76.5% (members of court) and 66.9% (general population). The allegations of the plaintiff in the negative declaratory action condition (the alleged debtor) were believed to a degree of 27.8% (members of court) and 41.1% (general population). Since the two propositions are incompatible and one must be true, the total in each condition should add up to 100%. Both for the general population (108%) and the members of court (104.3%), the total is slightly above 100%.

Next, the coefficient weights for a binary logistic regression with grant of request as the dependent variable and conviction (degree of belief on a scale from 0 to 100), membership of court (dummy variable, 1 = judge or judicial clerk), order of question (dummy variable, 1 = dichotomous decision before degree of belief) and type of action (dummy variable, 1 = "action" condition) were estimated (Model 1). The results are reported in Table 5. Unsurprisingly, the probability of grant (more precisely, the natural logarithm of the odds ratio $Pr(grant_i)/Pr(1\text{-}grant_i)$) increases with an increase in the conviction that the allegations supporting the claim are true. The order of questions and membership of court do not exert a significant influence, but the type of request does, with the negative declaratory action being granted at a much lower rate.

Table 5: Binary logistic regressions

	Model 1	Model 2	Model 3
[B]elief	0.081***	0.085***	0.087***
	(0.009)	(0.009)	(0.010)
[C]ourt	–0.291	–0.358	–0.463
	(0.320)	(0.330)	(0.353)
[O]rder	0.044	0.045	0.001
	(0.300)	(0.308)	(0.333)
[T]ype of request	1.290***	1.295***	1.193**
	(0.386)	(0.386)	(0.414)
[M]ale		0.368	0.425
		(0.315)	(0.347)
[A]ge		– 0.040**	–0.042**
		(0.013)	(0.015)
[L]oss aversion			0.007
			(0.157)
[L] x [T]			–0.796*
			(0.359)
Intercept	–6.295***	–5.039***	–4.225***
	(0.699)	(0.798)	(0.836)
Observations	383	383	337
Pseudo R^2	0.470	0.490	0.501

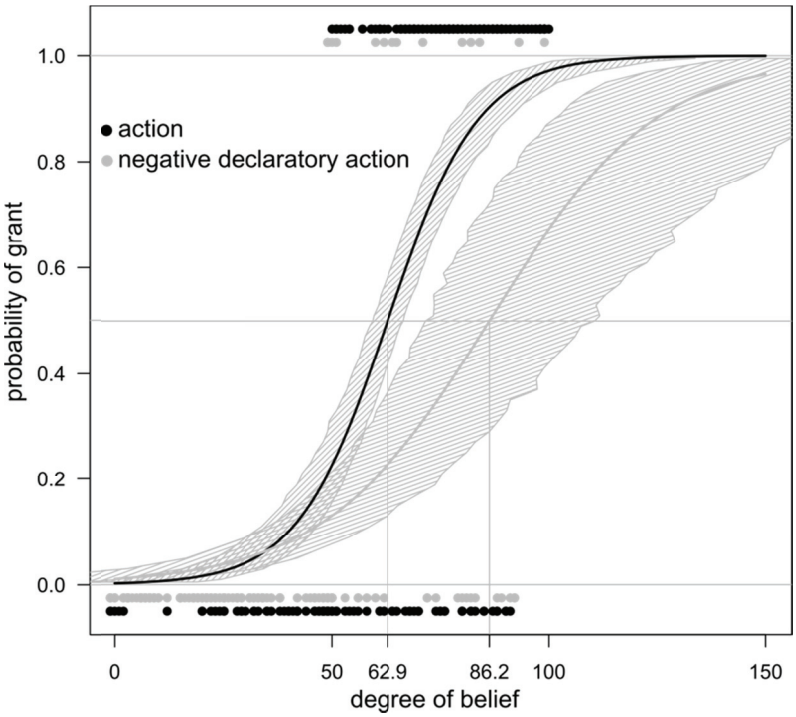
*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Continuing with the discussion of Model 1, the relationship between degree of belief and probability of grant is displayed graphically. Figure 2 plots the function of degree of belief on probability of grant for both types of claims. The grey shaded areas indicate the 95%-confidence intervals (bias corrected and accelerated), estimated by bootstrapping.[81] Since few participants granted the negative declaratory action, the confidence intervals for the negative declaratory action are much wider than for the action.

At a conviction of 62.9% (95%-confidence interval 59.4 – 66.5) the probability that the request in the action condition (order for payment) is granted is 50%. For the request in the negative declaratory action condition (declaration that nothing is due) to be granted with a probability of 50%, the degree of belief must be 86.2% (95%-confidence interval 71.6 – 100).

[81] 2,000 bootstrap replicates were generated using the R (R Core Team, *R: A language and environment for statistical computing*. http://www.R-project.org (01 July 2013)) command „boot ()" (A. Canty and B. Ripley, *boot: Bootstrap R (S-Plus) Functions: R package version 1.3-9*. http://cran.r-project.org/web/packages/boot/ (01 July 2013) implementing A. C. Davison and D. V. Hinkley, *Boostrap methods and their applications*, Cambridge series in statistical and probabilistic mathematics (New York: Cambridge University Press, 1997)). The 95% confidence interval was calculated as adjusted bootstrap percentile (Bca) as suggested by Wright, Strubler and Vallano, 'Statistical techniques for juror and jury research', 96 sq.

Figure 2: Plot of the function of degree of belief on probability of grant

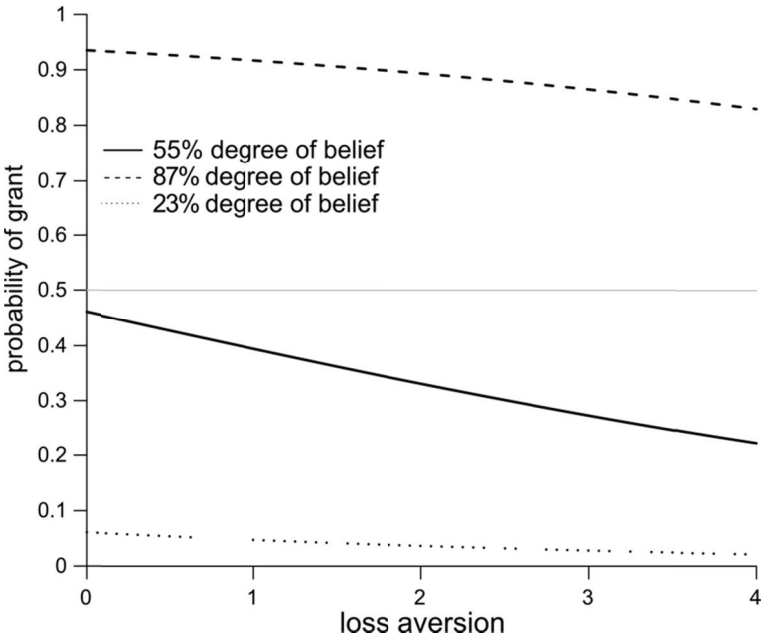


In a next step, Model 1 is expanded by including the age and gender of the participant as independent variables (Model 2). Gender has no influence on the probability of grant. Age has a negative influence, i.e., the older the participant, the less likely he or she is to grant the request.

In a final step, Model 2 is expanded by including loss aversion and the interaction of loss aversion and type of action (Model 3). Since it is generally assumed that non-monotonic valuations are incoherent, for the last model, all participants that exhibited non-monotonic valuations were excluded from the analysis. 5 (3%) of the court members and 41 (18%) of the general population sample showed non-monotonic valuations, which leaves 337 observations (152 members of court and 185 members of the general population). The interaction of loss aversion and type of request is significant. With increasing loss aversion, the probability of grant of the request in the action condition (order of payment) decreases.

Figure 3 plots the function of loss aversion on probability of grant in the action condition for a male, 40 year old member of court that first decides on the action and then states his conviction. The loss aversion is expressed by the Gain/Loss-ratio as explained in Section III D above, i.e., the ratio at which the participant is indifferent between a loss and gain with equal probabilities. The degree of belief is set at three levels: 55%, which is the mean in the action condition; 23%, one standard deviation below the mean; and 87%, one standard deviation above the mean. For a degree of belief of 55%, the probability that the action request is granted is about 47% for a participant exhibiting no loss aversion, but only about 25% for a participant exhibiting a G/L-ratio of 4 (indicating strong loss aversion).

Figure 4: Plot of the function of loss aversion on probability of grant for three fixed levels of conviction



VI. Discussion

The results reveal stark differences in the estimated decision threshold in civil cases depending on the measurement method. The direct rating method reveals a decision threshold of judges and judicial clerks of about 90%, which aligns with the quantification suggested in the German and Swiss legal literature and with results from an empirical study by *Berger-Steiner* with a sample of Swiss judges.[82] The results from the direct rating method also show that there is a small, but non-negligible, minority of judges that believe any decision granting a request is only permissible if there is a 100% certainty that the factual allegations supporting the claim are true.

A decision threshold of 90% implies that a false positive error (erroneously granting the request) has a disutility that is about 10 times that of a false negative decision (erroneously denying the request). However, when the error costs associated with each decision are elicited directly, judges and judicial clerks do not express error costs that would support the high default standard of proof of "full conviction" applicable in Continental European Civil Law jurisdictions such as Germany and Switzerland. They rather share the assumption underlying the "preponderance of the evidence" or "balance of probabilities" standard of proof of the Common Law. This makes it doubtful whether the Civil Law's higher standard of proof is defensible in a decision theoretic framework.

[82] Berger-Steiner, *Beweismass,* 221.

Finally, the degree of belief in the truth of the allegations supporting the request in the action condition at which the probability of grant is 50% is only 63%, while it is 86% for the request in the declaratory action condition. *Zamir* and *Ritov* have estimated the decision threshold of Israeli trial lawyers using the rank-order method at 70%.[83] Since Israel applies the Common Law's "preponderance of the evidence" standard in civil disputes, the decision threshold of the Israeli lawyers is arguably too high, while that of the Swiss judges and judicial clerks should be higher if it were to conform to the normatively required "full conviction". The results suggest that there may be a "natural" or "intuitive" decision threshold that is largely unaffected by the normative standard of the respective legal system. They also lend empirical support to the claim that *in fact*, rather than in theory, the standard of proof used by courts in Continental Europe in civil cases is not much different from the Common Law's standard. The heated debate between *Clermont/Sherwin* and *Taruffo* may be much ado about nothing.

No fully convincing explanation exists for the higher decision threshold observed in the negative declaratory action condition. It is certainly true that a negative declaratory action is a rare beast. As a general rule, it is up to the alleged creditor to decide when and whether to bring suit. Swiss law therefore requires that the plaintiff in a negative declaratory action shows a particular legal interest ("Rechtsschutzinteresse") in an immediate decision.[84] The scenario was silent whether the conditions for a negative declaratory action were met. This may explain why many judges were hesitant to grant the request. On the other hand, the sample of the general population that shows the same hesitation was most likely not aware of the legal requirements for a negative declaratory action.

The influence of age on the degree of belief required for a grant of the request was not expected. It is known that the probability of convicting an accused in a criminal trial increases with the age of the fact finder.[85] The effect of age in the civil case goes in the opposite direction, i.e., the higher the age, the lower the probability of grant. Neither for the effect of age on conviction in criminal cases nor for the effect on the grant in civil cases do convincing causal explanations that I am aware of exist.

The results further show that the degree of the individual's loss aversion influences the decision threshold in the hypothesized direction, i.e., a higher loss aversion leads to a higher decision threshold. While *Zamir* and *Ritov* suggest that loss aversion may provide a normative basis for a standard of proof in civil cases above 50%,[86] the results should not be interpreted to support this claim. While the status quo ante filing of the action is the *psychologically relevant* reference point, no good argument exists why it should be the *normatively*

---

[83] Zamir and Ritov, 'Loss Aversion, Omission Bias, and the Burden of Proof in Civil Litigation', 177.

[84] BGE 131 III 319 sect. 3.5; 120 II 20 sect. 3a.

[85] S. Anwar, P. Bayer and R. Hjalmarsson, 'A Fair and Impartial Jury? The Role of Age in Jury Selection and Trial Outcomes' (2012), NBER Working Paper.

[86] Zamir and Ritov, 'Loss Aversion, Omission Bias, and the Burden of Proof in Civil Litigation', 190.

relevant reference point. Why should not the status quo ante the act (or omission) that supports the legal claim be the relevant reference point? In the scenario used here, the creditor experiences a loss when he gave the loan to the debtor. The debtor experiences a loss when he is ordered to pay the loan back. Why is the loss aversion of the debtor taken into account (when justifying a decision threshold of above 50%), but not the loss aversion of the creditor? The same argument can be made for acts that lead to liability. The plaintiff whose car has been damaged by the negligent driving of the defendant may well experience the award of damages by the court as a gain, and the defendant as a loss, from the baseline "damaged car". But the plaintiff has certainly experienced the damage to his car as a loss from the (then relevant) baseline "intact car". Taking into account the loss aversion of the parties based on the status quo ante filing of the action seems to favour the defendant for no good reason.

A limitation of the present study is that it only used one scenario. It may be that different scenarios lead to different decision thresholds. Based on the current data, this cannot be ruled out. A further limitation is that the lottery choice task was not incentivized. This may have led to the relatively large proportion of the general population sample exhibiting non-monotonic valuations. In incentivized studies, typically about 10% of subjects show non-monotonic valuations.[87] Here, 18% of the general population sample, but only 4% of the members of court, exhibited an incoherent utility function.

VII. Conclusion

Judges from a Civil Law country (Switzerland) seem to use a much lower decision threshold in civil cases than required by legal doctrine and case law and indicated by themselves when asked directly. The actual difference in standards of proof in civil cases between Common Law and Civil Law may therefore be much smaller than thought. The error costs the Swiss judges and judicial clerks associate with each possible outcome of a civil action imply an error-cost-minimizing decision threshold of just 51%, the same as the threshold usually stated for the Common Law's "preponderance of the evidence" or "balance of probabilities" standard of proof. It could also be demonstrated that an individual's loss aversion as measured by the lottery choice task is predictive for his or her decision threshold, with individuals exhibiting stronger loss aversion having a higher decision threshold. Since the study is based on a single scenario, the external validity is limited.

[87] Holt and Laury, 'Risk Aversion and Incentive Effects', 1648; Gächter, Johnson and Herrmann, 'Individual-level loss aversion in riskless and risky choices', 13.

References

Anderson, T., Schum, D. A. and Twining, W. L., *Analysis of evidence*, Law in context, 2nd edn. (Cambridge: Cambridge Univ. Press, 2006).

Anwar, S., Bayer, P. and Hjalmarsson, R., 'A Fair and Impartial Jury? The Role of Age in Jury Selection and Trial Outcomes' (2012), *NBER Working Paper*.

Babcock, L., Farber, H. S., Fobian, C. and Shafir, E., 'Forming Beliefs About Adjudicated Outcomes: Perceptions of Risk and Reservation Values' (1995) 15, *International Review of Law and Economics*, 289–303.

Ball, V., 'The moment of truth: Probability theory and standards of proof' (1960) 14, *Vanderbilt Law Review*, 807–30.

Beattie, J. and Loomes, G., 'The Impact of Incentives Upon Risky Choice Experiments' (1997) 14, *Journal of Risk and Uncertainty*, 155-168.

Bender, R., 'Das Beweismaß', in W. Grunsky, R. Stürner, G. Walter and M. Wolf (eds.), *Festschrift für Fritz Baur* (Tübingen: Mohr Siebeck, 1981), pp. 247–71.

Berger-Steiner, I., *Das Beweismass im Privatrecht: Eine dogmatische Untersuchung mit Erkenntniswert für die Praxis und die Rechtsfigur der Wahrscheinlichkeitshaftung* (Bern: Stämpfli, 2008).

Bourmistrov-Jüttner, E., *Subjektive Wahrscheinlichkeitstheorie und rationale Entscheidungstheorie in Anwendung auf die Rechtspraxis* (München, 1987).

Brinkmann, M., *Das Beweismaß im Zivilprozess aus rechtsvergleichender Sicht* (Köln, Berlin, München: Heymann, 2005).

Bühler, A., 'Beweismass und Beweiswürdigung bei Gerichtsgutachten - unter Berücksichtigung der jüngsten Lehre und Rechtsprechung' (2010), *Jusletter 21 June 2010*.

Camerer, C. F. and Hogarth, R. M., 'The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework' (1999) 19, *Journal of Risk and Uncertainty*, 7-42.

Canty A. and Ripley B., *boot: Bootstrap R (S-Plus) Functions: R package version 1.3-9*. http://cran.r-project.org/web/packages/boot/ (1 July 2013).

Clermont, K. M., 'Standards of Proof Revisited' (2009) 33, *Vanderbilt Law Review*, 469–88.

Clermont, K. M. and Sherwin, E., 'A Comparative View of Standards of Proof' (2002) 50, *American Journal of Comparative Law*, 243–76.

Cullison, A. D., 'Probability Analysis of Judicial Fact-Finding: A Preliminary Outline of the Subjective Approach' (1969) 1, *University of Toledo Law Review*, 538–98.

Dane, F., 'In search of reasonable doubt' (1985) 9, *Law and Human Behavior*, 141–58.

Davison, A. C. and Hinkley, D. V., *Boostrap methods and their applications*, Cambridge series in statistical and probabilistic mathematics (New York: Cambridge University Press, 1997).

Deppenkemper, G., *Beweiswürdigung als Mittel prozessualer Wahrheitserkenntnis: Eine dogmengeschichtliche Studie zu Freiheit, Grenzen und revisionsgerichtlicher Kontrolle tatrichterlicher Überzeugungsbildung (§ 261 StPO, § 286 ZPO)* (Göttingen: V&R Unipress, 2004).

Dhami, M. K., 'On measuring quantitative interpretations of reasonable doubt' (2008) 14, *Journal of Experimental Psychology: Applied*, 353–63.

Einmahl, M., 'Zeugenirrtum und Beweismaß im Zivilprozess: Eine Fallstudie am Beispiel des Verkehrsunfallprozesses' (2001), *NJW*, 469–75.

Engel, C., 'Preponderance of the Evidence versus intime conviction: A Behavioral Perspective on a Conflict between American and Continental European Law' (2009) 33, *Vermont Law Review*, 435–67.

Etchart-Vincent, N. and l'Haridon, O., 'Monetary incentives in the loss domain and behavior toward risk: An experimental comparison of three reward schemes including real losses' (2011) 42, *Journal of Risk and Uncertainty*, 61-83.

Fehr, E. and Goette, L., 'Do Workers Work More if Wages Are High? Evidence from a Randomized Field Experiment' (2007) 97, *American Economic Review*, 298–317.

Fuchs, B., *Das Beweismaß im Arzthaftungsprozess* (Frankfurt am Main: Peter Lang, 2005).

Gächter, S., Johnson, E. J. and Herrmann, A., 'Individual-level loss aversion in riskless and risky choices' (2010), *CeDEX discussion paper series.*

Gottwald, P., 'Das flexible Beweismaß im englischen und deutschen Zivilprozess', in P. Gottwald, E. Jayme and D. Schwab (eds.), *Festschrift für Dieter Henrich zum 70. Geburtstag, 1. Dezember 2000* (Bielefeld: Gieseking, 2000), pp. 165–76.

Gräns, M., *Das Risiko materiell fehlerhafter Urteile: (Diss. Uppsala 1995)* (Berlin: Duncker & Humblot, 2002).

Greger, R., *Beweis und Wahrscheinlichkeit: Das Beweiskriterium im Allgemeinen und bei den sogenannten Beweiserleichterungen* (Köln: Heymann, 1978).

Guthrie, C., Rachlinski, J. J. and Wistrich, A. J., 'Inside the Judicial Mind' (2001) 86, *Cornell Law Review*, 777–830.

Harrison, G. W. and Rutström, E. E., 'Risk aversion in the laboratory', in J. C. Cox and G. W. Harrison (eds.), *Risk aversion in experiments* (Bingley: JAI, 2008), pp. 41–196.

Hastie, R., 'Algebraic models of juror decision processes', in R. Hastie (ed.), *Inside the Juror: The Psychology of Juror Decision Making* (Cambridge: Cambridge Univ. Press, 1993), pp. 84–115.

Holt, C. A. and Laury, S. K., 'Risk Aversion and Incentive Effects' (2002) 92, *American Economic Review*, 1644–55.

Hoyer, A., 'Der Konflikt zwischen richterlicher Beweiswürdigungsfreiheit und dem Prinzip "in dubio pro reo"' (1993) 105, *Zeitschrift für die gesamte Strafrechtswissenschaft*, 523–56.

Kadner Graziano, T., '"Alles oder nichts" oder anteilige Haftung bei Verursachungszweifeln?: Zur Haftung für perte d'unce chance/loss of chance und eine Alternative' (2011) 19, *Zeitschrift für Europäisches Recht*, 171–200.

Kahneman, D. and Tversky, A., 'Prospect Theory: An Analysis of Decision under Risk' (1979) 47, *Econometrica*, 263–92.

Kaplan, J., 'Decision theory and the factfinding process' (1968) 20, *Stanford Law Review*, 1065–92.

Kaye, D. H., 'The Limits of the Preponderance of the Evidence Standard: Justifiable Naked Statistical Evidence and Multiple Causation' (1982) 1982, *Law & Social Inquiry*, 487–516.

Kaye, D. H., 'Apples and Oranges: Confidence Coefficients and the Burden of Persuasion' (1987) 73, *Cornell Law Review*, 4–77.

Kaye, D. H., 'Bayes, Burdens and Base Rates' (2000) 4, *International Journal of Evidence and Proof*, 260–7.

Kaye, D. H., 'The error of equal error rates' (2002) 1, *Law, Probability & Risk*, 3–8.

Kegel, G., 'Der Individualanscheinsbeweis und die Verteilung der Beweislast nach überwiegender Wahrscheinlichkeit', in K. H. Biedenkopf, H. Coing and E.-J. Mestmäcker (eds.), *Das Unternehmen in der Rechtsordnung: Festgabe für Heinrich Kronstein aus Anlass seines 70. Geburtstages am 12. September 1967* (Karlsruhe: C.F. Müller, 1967), pp. 321–44.

Kinsch, P., 'Probabilité et certitude dans la preuve en justice', in Institut Grand-Ducal (ed.), *Actes de la Section des Sciences Morales et Politiques* (Luxemburg, 2009), pp. 65–103.

Köbberling, V. and Wakker, P. P., 'An index of loss aversion' (2005) 122, *Journal of Economic Theory*, 119–31.

Korobkin, R. and Guthrie, C., 'Psychological Barriers to Litigation Settlement: An Experimental Approach' (1994) 93, *Michigan Law Review*, 107–92.

Laudan, L. and Saunders, H. D., 'Re-Thinking the Criminal Standard of Proof: Seeking Consensus about the Utilities of Trial Outcomes' (2009) 7, *International Commentary on Evidence*, 1–36.

Lee, T. R., 'Pleading and Proof: The Economics of Legal Burdens' (1997), *Brigham Young University Law Review*, 1–34.

Lempert, R. O., 'Modeling Relevance' (1977) 75, *Michigan Law Review*, 1021–57.

Lillquist, E., 'Recasting Reasonable Doubt: Decision Theory and the Virtues of Variability' (2002) 36, *U.C. Davis Law Review*, 85–197.

Maassen, B. M., *Beweismaßprobleme im Schadenersatzprozess* (Köln: Carl Heymanns Verlag KG, 1975).

McBride, E., 'Is the civil 'higher standard of proof ' a coherent concept?' (2009) 8, *Law, Probability & Risk*, 323–51.

Motsch, R., 'Vom Prozess als Beweis zum Überwiegensprinzip', in U. Klug, T. Ramm, F. Rittner and B. Schmiedel (eds.), *Gesetzgebungstheorie, juristische Logik, Zivil- und Prozessrecht: Gedächtnisschrift für Jürgen Rödig* (Berlin, New York: Springer, 1978), pp. 334–48.

Motsch, R., 'Comparative and Analytical Remarks on Judicial Fact-Finding', in H. Rüßmann (ed.), *Festschrift für Gerhard Käfer* (Saarbrücken: Juris GmbH, 2009), pp. 241–71.

Nagel, S., Lamm, D. and Neef, M., 'Decision theory and juror decision making', in B. D. Sales (ed.), *The trial process*, Perspectives in law and psychology (New York: Plenum Press, 1981), vol. 2, pp. 353–86.

Nell, E. L., *Wahrscheinlichkeitsurteile in juristischen Entscheidungen* (Berlin: Duncker & Humblot, 1983).

O'Malley, K. F., Grenig, J. E. and Lee, W. C. (eds.), *Federal jury practice and instructions* (Eagan, MN: Westlaw, 2001).

Posner, R. A., 'An Economic Approach to the Law of Evidence' (1999) 51, *Stanford Law Review*, 1477–546.

Prelec, D., 'The Probability Weighting Function' (1998) 66, *Econometrica*, 497–527.

Prütting, H., '§ 294', in T. Rauscher, P. Wax and J. Wenzel (eds.), *Münchener Kommentar zur Zivilprozessordnung*, 3.th edn. (München: Beck, 2008).

Prütting, H., 'Beweislast und Beweismaß: Der Einfluss Leo Rosenbergs und Karl Heinz Schwabs auf die Entwicklung eines modernen Beweisrechts' (2010) 123, *Zeitschrift für Zivilprozess*, 135–45.

R Core Team, *R: A language and environment for statistical computing*. http://www.R-project.org (1 July 2013).

Rabin, M., 'Risk Aversion and Expected-utility Theory: A Calibration Theorem' (2000) 68, *Econometrica*, 1281–92.

Rachlinski, J. J., 'Gains, Losses and the Psychology of Litigation' (1996), *Southern California Law Review*, 113–85.

Rechberger, W. H., 'Maß für Maß im Zivilprozess?: Ein Beitrag zur Beweismaßdiskussion', in H. Prütting (ed.), *Festschrift für Gottfried Baumgärtel zum 70. Geburtstag* (Köln: Carl Heymanns Verlag KG, 1990), pp. 471–90.

Redmayne, M., 'Standards of Proof in Civil Litigation' (1999) 62, *Modern Law Review*, 167–95.

Sand, L. B., Rakoff, J. S., Reiss, S. A., Loughlin, W. P., Allen, S. W. and Siffert, J. S. (eds.), *Modern Federal Jury Instructions - Civil Volumes* (Los Angeles: LexisNexis, 2007).

Sheppard, S., 'The Metamorphoses of Reasonable Doubt: How Changes in the Burden of Proof Have Weakened the Presumption of Innocence' (2003) 78, *Notre Dame Law Review*, 1165–250.

Simon, R. J., '"Beyond a Reasonable Doubt": An Experimental Attempt at Quantification' (1970) 6, *Journal of Applied Behavioral Science*, 203–9.

Simon, R. J. and Mahan, L., 'Quantifying Burdens of Proof: A View from the Bench, the Jury, and the Classroom' (1971) 5, *Law & Society Review*, 319–30.

Stein, A., *Foundations of evidence law* (Oxford, New York: Oxford Univ. Press, 2005).

Summermatter, D. and Jacober, C., 'Zum Beweismass beim Kausal- und Motivationszusammenhang: Versuch einer begrifflichen Klärung anhand der jüngeren Rechtsprechung im Zivil-, Straf- und Sozialversicherungsrecht' (2012), *HAVE*, 136–49.

Taruffo, M., 'Rethinking the Standards of Proof' (2003) 51, *American Journal of Comparative Law*, 659–77.

Tillers, P. and Gottfried, J., 'Case comment--United States v. Copeland, 369 F. Supp. 2d 275 (E.D.N.Y. 2005): A Collateral Attack on the Legal Maxim That Proof Beyond A Reasonable Doubt Is Unquantifiable?' (2007) 5, *Law, Probability & Risk*, 135–57.

Tribe, L. H., 'Trial by Mathematics: Precision and Ritual in the Legal Process' (1971) 84, *Harvard Law Review*, 1329–93.

Tversky, A. and Kahneman, D., 'Loss Aversion in Riskless Choice: A Reference-Dependent Model' (1991) 106, *Quarterly Journal of Economics*, 1039–61.

Tversky, A. and Kahneman, D., 'Advances in prospect theory: Cumulative representation of uncertainty' (1992) 5, *Journal of Risk and Uncertainty*, 297–323.

Tyree, A. L., 'Proof and probability in the Anglo-American legal system' (1982) 23, *Jurimetrics*, 89–100.

Wagner, G., 'Ökonomische Analyse der Rechtsmittel', in R. Bork, T. Eger and H.-B. Schäfer (eds.), *Ökonomische Analyse des Verfahrensrechts: Beiträge zum XI. Travemünder Symposium zur Ökonomischen Analyse des Rechts (26. bis 29. März 2008)* (Tübingen: Mohr Siebeck, 2009), pp. 156–93.

Walter, G., *Freie Beweiswürdigung* (Tübingen: Mohr Siebeck, 1979).

Walter, H. P., 'Beweis und Beweislast im Haftpflichtprozess' (2009), *Haftpflichtprozess*, 47–68.

Wright, D. B., Strubler, K. A. and Vallano, J. P., 'Statistical techniques for juror and jury research' (2011) 16, *Legal and Criminological Psychology*, 90–125.

Wright, R. W., 'Proving Facts: Belief versus Probability', in H. Koziol and B. C. Steininger (eds.), *Tort and Insurance Law* (Vienna: Springer Vienna, 2009), pp. 79–105.

Zamir, E. and Ritov, I., 'Loss Aversion, Omission Bias, and the Burden of Proof in Civil Litigation' (2012) 41, *Journal of Legal Studies*, 165–207.

Appendix I

Imagine that you are a judge that has to decide the following case.

Mr Arnold, the plaintiff, has sued Mr Graf, the defendant, for payment of CHF [EUR] 20,000. *[Note: in the "negative declaratory action" condition, the preceding sentence was replaced by: "Mr Graf, the plaintiff, wants a declaration that he does not owe Mr Arnold, the defendant, any money."]*

Mr Arnold claims that he lent Mr Graf CHF [EUR] 20,000 and Mr Graf had not paid him back. Mr Graf denies ever having received a loan from Mr Arnold.

It is undisputed that Mr Arnold and Mr Graf have known each other for years, since they attended college together, and have a long running business relationship. Mr Arnold is a business lawyer, Mr Graf business man running his own business.

According to Mr Arnold, Mr Graf asked him for a loan of CHF [EUR] 30,000 for a couple of months to overcome financial difficulties. Mr Arnold agreed to lend Mr Graf CHF [EUR] 20,000. Because the two knew each other for so long, he did not insist on a written contract. According to Mr Arnold's testimony, he gave the money in cash to Mr Graf, without getting a receipt, on 1 December 2010 (roughly one year before filing suit).

A bank statement of Mr Graf's account shows a deposit of CHF [EUR] 20,000 on 1 December 2010.

Mr Graf testifies that he had indeed asked Mr Arnold for a loan, and Mr Arnold initially agreed to lend him CHF [EUR] 20,000, but then changed his mind and did not give him any money. He, i.e., Mr Graf, had gotten a loan from another party to overcome his financial difficulties. He would not identify the third party.

A witness called by Mr Arnold testifies that Mr Graf told her on 2 December 2010 that he had received a loan from Mr Arnold in the amount of CHF [EUR] 20,000 on 2 December 2010. Mr Graf remarks that he cannot remember having had a conversation with the witness.

The wife of Mr Arnold testifies that her husband had told her at the end of November 2010 that he intended to loan CHF [EUR] 20,000 to Mr Graf, for tax reasons in cash, without a receipt.

Mr Graf says that it is inconceivable that Mr Arnold, a business lawyer, would give him such a large amount of money without a written contract and without demanding a receipt.

Note: The scenario was in German. In Germany, the amount was given in Euro, in Switzerland, in Swiss Franc. One Swiss Franc at the time of the survey was approximately EUR 0.9. Respondents could access the full text of the scenario while answering the questions using a help button.

Appendix II

Table 6: Descriptive statistics for the degree of belief and decision without exclusion of 24 subjects

	Action				negative declaratory action			
	grant		denial		grant		denial	
	obs. (share)	Ø conviction (SD)	obs. (share)	Ø conviction (SD)	obs. (share)	Ø conviction (SD)	obs. (share)	Ø conviction (SD)
members of court (N = 160)	63 (72%)	83.0 (13.5)	25 (29%)	57.7 (19.0)	6 (8%)	65.2 (19.0)	66 (92%)	22.4 (23.0)
general population (N = 247)	99 (59%)	77.8 (15.7)	70 (41%)	50.6 (24.7)	24 (30%)	51.6 (20.8)	54 (70%)	37.9 (27.3)

Table 7: Binary logistic regressions without exclusion of 24 subjects

	Model 1	Model 2	Model 3
[B]elief	0.055***	0.057***	0.062***
	(0.006)	(0.006)	(0.007)
[C]ourt	–0.253	–0.299	–0.480
	(0.269)	(0.257)	(0.302)
[O]rder	-0.008	0.003	0.008
	(0.250)	(0.259)	(0.286)
[T]ype of request	0.652*	0.651*	0.601
	(0.302)	(0.306)	(0.342)
[M]ale		0.312	0.573
		(0.263)	(0.297)
[A]ge		– 0.035**	–0.044***
		(0.011)	(0.013)
[L]oss aversion			0.076
			(0.132)
[L] x [T]			–0.982***
			(0.286)
Intercept	–3.772***	–2.612***	–2.177***
	(0.425)	(0.561)	(0.640)
Observations	407	407	361
Pseudo R^2	0.323	0.344	0.379

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$